Genetic Influences on Health
Does Race Matter?

Mike Bamshad, MD

Race is frequently used by clinicians and biomedical researchers to make inferences about an individual's ancestry and to predict whether an individual carries specific genetic risk factors that influence health. The extent to which race is useful for making such predictions depends on how well race corresponds with genetic inferences of ancestry, how frequently common diseases in different racial groups are influenced by the same vs different gene variants, and whether such variants have the same effects in different racial groups. New studies of human genetic variation show that while genetic ancestry is highly correlated with geographic ancestry, its correlation with race is modest. Therefore, while data on the correspondence of race, ancestry, and health-related traits are still limited, particularly in minority populations, geographic ancestry and explicit genetic information are alternatives to race that appear to be more accurate predictors of genetic risk factors that influence health. Making accurate ancestry inferences is crucial because common diseases and drug responses are sometimes influenced by gene variants that vary in frequency or differ altogether among racial groups. Thus, operationalizing alternatives to race for clinicians will be an important step toward providing more personalized health care.

JAMA. 2005;294:937-946 www.jama.com

An important goal of 21st-century medicine is to predict an individual's medical future—that is, to identify the set of risk factors for disease and predictors of treatment response that influence a person's health, with the goal of more effectively treating and preventing disease. Throughout much of the world, race—using its historical meaning as a descriptor of Africans, Asians, Europeans, Native Americans, and Pacific Islanders—is often considered one key determinant of health.[1,2] Race might influence an individual's health in several ways. It might covary with different environmental or genetic factors that underlie risk, different interactions between genetic and environmental factors, or different combinations thereof. Many environmental factors that influence health are known, but most of the genetic factors, much less interactions between the two, remain to be discovered. Nevertheless, there is widespread speculation that genetic factors influencing health differ among racial groups because many health-related traits vary among racial groups.[3] This speculation has revived a long-standing debate in medical and scientific communities about the validity and necessity of using race to make inferences about an individual's genetic ancestry, some scientists embracing this idea and others dismissing it.[4-8]

In this debate, one issue that is commonly confused is the difference between race and ancestry. *Ancestry* refers to objective genetic relationships between individuals and among populations, whereas *race* has always been a somewhat arbitrary definition of population boundaries. For example, while an individual might have ancestors from Europe, Africa, and North America, he or she still might be categorized as an African American. Therefore, race captures some biological information about ancestry, but it is not equivalent to ancestry. Yet clinicians often want to know whether it is valid and reliable to use race as a proxy to infer an individual's genetic risk for disease and treatment response. Whether race matters is, however, complicated because it depends on the relationship between the genetic risk factor, ancestry, and race. For example, β-blockers and angiotensin-converting enzyme (ACE) inhibitors for hypertension treatment may not work as well, on average, in African Americans compared with European Americans, but both types of drugs appear to work perfectly well in a large fraction of African Americans and poorly in some European Americans.[9] The observation might be explained by a hypothetical genetic predictor of positive response to ACE inhibitors that is common in European Americans but that is also present in some African Americans because of admixture and absent in some

Author Affiliation: Department of Pediatrics and Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah, Salt Lake City.
Corresponding Author: Mike Bamshad, MD, Department of Pediatrics and Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT 84112 (mike@genetics.utah.edu).

European Americans (FIGURE 1). In such a case, the best predictor of treatment response might be the presence of the variant (ie, direct testing); the next best might be an accurate estimator of genetic ancestry, and race might be only a poor predictor of genetic risk and therefore treatment response.

Two additional related but separate issues—varied genetic causes of heath-related traits among racial groups vs the influence of genetic factors on health-related traits that vary in prevalence among racial groups—have also frequently been conflated with one another. For example, genetic variants putatively associated with hypertension,[10,11] diabetes,[12-14] atherosclerosis,[15] and many autoimmune disorders[16] are common in only a single racial group or differ significantly in frequency among groups. Therefore, disease risk and treatment response clearly are, in some circumstances, influenced by genetic factors or genetic effects that vary among racial groups.[17-21] Yet, it is unclear whether such genetic risk factors explain, even partly, variation in the prevalence of these diseases among racial groups. Indeed, many health-related disparities probably are only modestly affected by genetics, influenced more strongly instead by environmental factors such as dietary differences and inequities in the provision of health care services.[22-24]

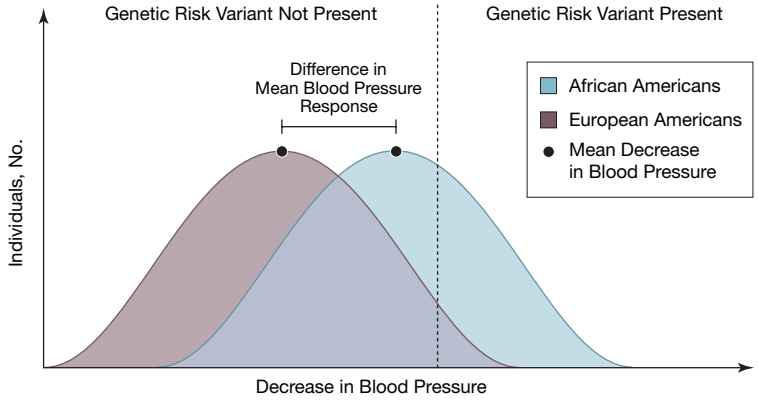
The extent to which race is useful for making predictions about genetic differences that influence health can be informed by considering several questions: (1) How well do "traditional" classifications of race correspond with genetic inferences of individual ancestry? (2) How frequently are common diseases in different racial groups influenced by the same vs different gene variants? (3) Do gene variants associated with common diseases have the same effects in different racial groups? While none of these questions can be neatly resolved, new data on patterns of human genetic variation and gene variants influencing health provide guidance in a field where strong opposing opinions are common.

Correspondence Between Race and Population Structure

Two randomly chosen humans differ at approximately 1 in 1000 nucleotide pairs—dubbed single nucleotide polymorphisms (SNPs)—so, on average, 2 humans differ at approximately 3 million of the genome's 3 billion nucleotides (ie, they are 99.9% identical).[25] Collectively, it is estimated that there are approximately 10 million SNPs with a frequency of at least 1% and millions more that are vanishingly rare.[26] While some of these SNPs contribute to phenotypic variation, most of this variation is said to be "neutral" or without functional consequence.[27] The distribution of this neutral variation reflects human demographic history including the organization of humans into subpopulations (ie, population structure).[28,29]

Studies using a broad range of genetic markers have confirmed that approximately 10% to 15% of total genetic variation in humans is explained by differences between sub-Saharan Africans, Northern Europeans, and East Asians.[30-33] This means that individuals from different populations are, on average, slightly more different from one another than individuals from the same population (FIGURE 2A,B). Yet, while the fraction of overall genetic variation distributed between groups is relatively small, it is highly structured (allele frequencies at different loci are highly correlated)[34] so that a modest number of genotypes (ie, several hundred per person) can, with a high degree of accuracy, allocate anonymous individuals to groups that correspond to ancestry from different geographic regions (Figure 2C).[30-32,35] Thus, geographic ancestry can be used to make reasonably accurate predictions of genetic ancestry, but it is not perfect. Populations from neighboring geographic regions typically share more recent common ancestors, and therefore their allele frequencies are more highly correlated—a pattern commonly manifest as a cline of allele frequencies.[36] Because of such clines, individuals sampled continuously across some large geographic regions (eg, Middle East, Central Asia) are difficult to allocate into genetic groups that are inclusive of all individuals from these regions. Correspondence with geography is also less apparent for populations (eg, Hispanics, South Asians) that are historically admixed.

Figure 1. Hypothetical Relationship Between Genetic Risk, Ancestry, and Race

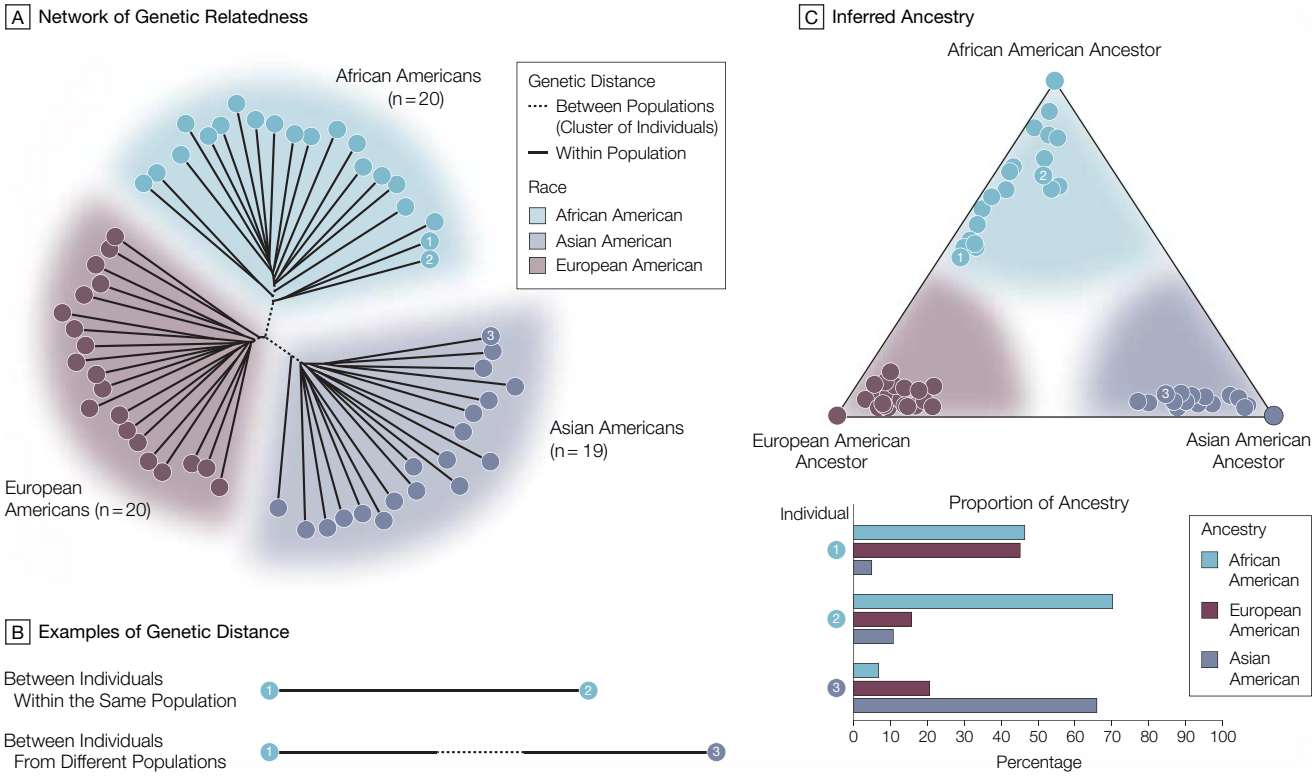


Distributions of the reduction in blood pressure observed in African Americans and European Americans after treatment with an angiotensin-converting enzyme (ACE) inhibitor. One hypothetical explanation for the mean difference in treatment response is that a genetic risk variant predictive of a positive response to treatment is more common in European Americans (individuals to the right of the dotted line) than in African Americans. Note, however, that some African Americans also have the genetic risk variant and that many African Americans and European Americans who do not have the genetic risk variant have a similar response to treatment (ie, overlap between distributions). In this case, race might not be considered a good predictor of genetic risk or response to treatment. Based on original concept by Seghal.[9]

What then is the best way to represent the fraction of overall genetic variation that corresponds with geography, and how important is this variation to health-related traits? Some scientists advocate that the relatively crude racial categories proposed by the Office of Management and Budget and adopted by many federal agencies including the National Institutes of Health and the Food and Drug Administration continue to be used as proxies for geographic ancestry.[4,37] Others suggest that inference of group membership should, instead, be based on explicit genetic data, ignoring a priori racial proxies.[32,38,39] In the United States, several studies have reported that classification of individuals by self-identified racial group was highly correlated with inferences based on explicit genetic data. In the largest of these studies, Tang et al[37] genotyped 366 microsatellites in 3636 US subjects who self-identified their ancestry as European, African, Hispanic, or Asian. Group membership inferred from genetic data differed from self-identified race in only 5 individuals. Can these results be extrapolated across other study populations in the United States? Immigration to the United States has been relatively recent (≤10 generations), and the geographic ancestry of the majority of North Americans traces to only a few circumscribed regions of the world: European Americans to Western Europe, African Americans to West Africa,[40] and most Asian Americans to East Asia, particularly China.[41] The results of the study by Tang et al might be applicable across the United States, but this will need to be demonstrated empirically.

If some individuals can be sorted broadly into genetic groups concor-

Figure 2. Inference of Individual Ancestry Proportions From Genetic Data



(A) A network depicting the genetic relatedness among individuals (circles) with self-identified ancestry from Africa (20), Asia (19), and Europe (20) genotyped for 250 coding single nucleotide polymorphisms (SNPs) for which the less common allele has a frequency of at least 10% (Bamshad et al[32]). The length of each branch (black lines) is proportional to genetic distance between individuals and populations. Distinguishing individuals by race (shaded areas) obscures this variation in ancestry. The distance between any 2 circles of the same color (solid lines) is large and reflects high within-population variance, whereas the distance between clusters (dotted lines) is small and reflects low between-population variance. Individuals with a higher proportion of ancestry from more than one population (individual 2) are connected directly to the branches between clusters. (B) The genetic distance between individuals is reflected by the sum of the branch lengths between individuals. The genetic distance between individuals from different populations, such as individuals 1 and 3, is slightly greater than the genetic distance between individuals within the same population, such as individuals 1 and 2. Thus, despite the high within-population variance, individuals from different populations are, on average, more different from one another than individuals from the same population. (C) Inferred ancestry proportions for individuals (circles) used in panel A genotyped for 500 coding SNPs with a minor allele frequency of at least 10%. The distance of each circle to an apex is proportional to the degree of ancestry in African Americans, Asian Americans, or European Americans. The degree to which the circles are clustered within a population reflects the degree of admixture and structure within a population. The circles representing African Americans are less tightly clustered because the proportion of ancestry among individuals is more varied than in Asian Americans and European Americans. Distinguishing individuals by race (shaded areas) obscures this variation in ancestry. Data source for panels B and C, Bamshad et al.[32]

dant with their self-assessed racial identity, is race a "good" predictor of individual ancestry? The fraction of variation that an African American shares with West Africans varies considerably because African Americans have admixed to variable degrees with groups originating from other geographic regions. The West African contribution to individual African American ancestry averages about 80%, but ranges from approximately 20% to 100%.[42] The genetic composition of self-identified European Americans also varies, with approximately 30% of European Americans estimated to have less than 90% European ancestry (Figure 2C).[42] Accordingly, membership in a genetically inferred group does not mean that all members of the group necessarily have a similar genetic composition. For example, Reiner et al[43] found substantial admixture and population structure among 810 African Americans who were part of the Cardiovascular Health Study. Similarly, Hispanics from different regions of the United States are variably admixed with different populations (eg, more African admixture in Hispanics living in the Southeast vs more admixture with Native Americans in the Southwest). Knowing the proportion of an individual's ancestry that originated in different populations and to what degree a group is divided into genetic subpopulations can be useful for identifying genetic and environmental factors—by reducing false-negative associations and uncovering true associations—that underlie common diseases for which risk varies among populations.[4] To this end, several hundred loci that are particularly informative for estimating ancestry proportions in African, European, Asian, Hispanic, and Native Americans have been identified.[44]

In contrast to the situation in the United States, the geographic origins of individuals with African or Asian ancestry living in other parts of the world (eg, Europe, South America) are more heterogeneous,[45] which is important because sub-Saharan Africans and Asians are clearly divided into multiple genetic subpopulations.[30,31,46-48] Indeed, some populations from East Africa and West Africa are, on average, more different genetically than populations from Northern Europe and East Asia.[48,49] Moreover, some genetic variants are associated with differences in disease susceptibility among different African populations.[50] Outside the United States, groups defined by racial categories might therefore exhibit even more internal genetic differentiation.[51] Failure to account for this population structure could confound efforts to identify predictors of disease risk and treatment outcome. Accounting for such structure is more straightforward with genetic-based inferences but hard to represent with racial labels.

Worldwide, notions of race capture only a modest amount of information about geographic ancestry and therefore population structure and capture even less, in general, than ancestry inferences from explicit genetic data. Race will become an even more inaccurate proxy for ancestry as populations become increasingly admixed. Further caution is warranted because inferences about the correspondence between race, genetics, and geography have, to date, been extrapolated from a relatively small number of the world's populations and sampled from a limited number of geographic regions. The range of parameters under which the geographic origins of an anonymous sample can be identified is still poorly understood. This uncertainty reflects the need for an unbiased sampling of variation (eg, through resequencing) across the genome from individuals in well-characterized communities sampled from contiguous geographic regions throughout the world. Several private and public initiatives to sample human genetic variation more comprehensively,[52] most notably the International Human HapMap Project spearheaded by the National Human Genome Research Institute,[53] will provide valuable new data for making genetic inferences of geographic ancestry.

Genetic Risk Factors for Common Diseases Among Populations

To accurately assess the influence of genetic differences among individuals with different geographic ancestries on health-related traits, it is necessary to understand the possible causes of differences in genetic risk factors among populations. Differences in genetic risk among groups may be caused by (1) susceptibility variants that are present in one group but absent in others, (2) susceptibility variants that vary in frequency among groups, or (3) a variant that influences susceptibility in one population that might not have the same effect in a different population. The most direct way to study whether genetic risk factors vary among racial groups is to find susceptibility variants influencing health and determine whether these variants differ in frequency and/or effect among groups.

Most health-related traits, such as susceptibility to diabetes, obesity, infection, and cancer, are complex traits influenced by the combined effects of several or more gene variants, each with a modest effect, together with the environment. What does population genetics theory tell us to expect about the proportion of genetic risk factors for health shared among groups? A popular model of the genetic architecture of common disease, the "common disease/common variant hypothesis," predicts that most gene variants for complex diseases are common (ie, ≥10%) and therefore old and found in multiple groups rather than rare and population-specific.[54] Therefore, how often genetic influences on health-related traits differ between populations partly depends on the proportion of common disease—gene variants causing mendelian disorders often vary substantially among populations but these explain only a small proportion of health-related traits[55]—explained by rare vs common gene variants, an empirical question about which some data are available.

Our knowledge of true associations between a gene variant and risk of com-

mon disease or drug response in different populations is limited to several dozen "valid" gene-disease associations.[56,57] Many gene variants putatively associated with complex diseases are shared among populations. Yet variants influencing risk for many diseases including atherosclerosis,[14,15] hypertension,[9,10] sudden cardiac death,[58] asthma,[59] diabetes,[11-13] macular degeneration,[60] and common infections, including AIDS,[61,62] are common in only a single population or differ significantly in frequency among groups. For example, the 825T allele of *GNB3* (encoding the G protein β3 subunit) associated with both obesity and hypertension varies in frequency from approximately 80% in Africans, approximately 45% in Asians, and approximately 30% in Europeans.[63] Similarly, the frequency of the Glu23Lys type 2 diabetes risk allele of *KCNJ11* (encoding a potassium inwardly rectifying channel, subfamily J, member 11) varies from approximately 20% in Caribbean individuals of African ancestry to approximately 40% in European Americans.[12,13] The case appears to be similar for many gene variants underlying variable drug responses. Tate and Goldstein[38] recently found that of 42 gene variants associated with drug responses in 2 or more studies, more than two thirds showed a significant difference in frequency between African and European Americans. These differences could be of substantial clinical significance.

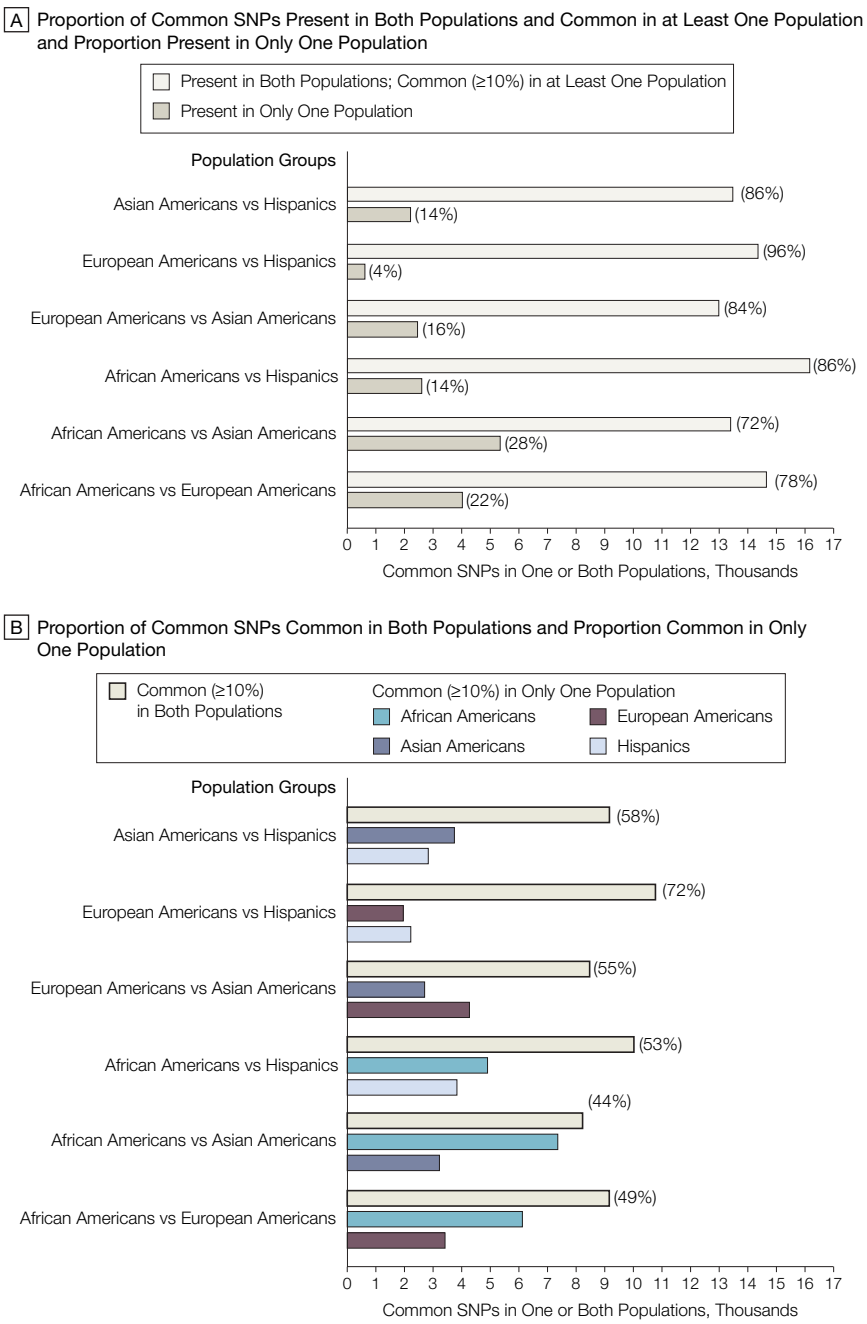
For other common diseases, susceptibility in different populations appears to be determined, in part, by risk variants in different genes. For example, 3 variants in *CARD15* (formerly *NOD2*)—R702W, G908R, and 1007fs—have been associated with Crohn disease in European Americans.[64] Also in populations of European ancestry, the variant R620W in *PTPN22* (encoding a protein tyrosine phosphatase) has been associated with susceptibility to several autoimmune disorders including rheumatoid arthritis[65] and systemic lupus erythematosus.[66] No variants in *CARD15* or *PTPN22* have yet been associated with these disorders in populations of either African or Asian ancestry.[67,68] The frequencies of many functional gene variants and most population-specific risk variants, such as those in *CARD15* and *PTPN22*, are often less than 10%. These findings suggest that regardless of whether the common disease/common variant model is correct only sometimes, as appears to be the case, inference of individual ancestry, by proxy or inference via explicit genetic data, will provide some information about the genetic factors influencing common health-related traits—and therefore might influence clinical management.

Variation in gene copy number between populations can also influence differences in disease risk among populations. The mean copy number of a segmental duplication on chromosome 17 encompassing the gene encoding CCL3L1 (MIP-1αP), a ligand for the human immunodeficiency virus (HIV) co-receptor, CC chemokine receptor 5 (CCR5), and a potent HIV-1 suppressive chemokine, ranges from approximately 2 in Europeans to 6 in Africans, with most Asians having about 3 to 4 copies.[69] Lower *CCL3L1* copy number is associated with lower *CCL3L1* levels, a higher viral set point, a faster decline in CD4 T cells, and increased HIV/AIDS susceptibility. However, it is not absolute *CCL3L1* copy number per se, but *CCL3L1* copy number standardized by the population mean *CCL3L1* copy number, that is associated with risk. Therefore, *CCL3L1* gene dose can be interpreted as a risk factor only when considered along with an individual's group membership or geographic ancestry, possibly because mean copy number is associated with other genetic factors that influence risk.

Are these examples likely to represent the situation for the genetic factors influencing most common diseases? Gene variants associated with complex diseases appear to cause molecular effects on protein function that are similar to those caused by normal genetic variation (ie, they both have modest effects on protein function compared with mutations causing mendelian disorders).[70] Accordingly, inspection of the frequency distribution of normal gene variants among populations of different geographic ancestry might inform us about how often to expect common risk variants for health-related traits to vary among groups. Several analyses of allele sharing among groups have suggested that most common gene variants are shared among African Americans, European Americans, and Asian Americans.[52,71,72] In most of these studies, alleles were measured by genotyping variants ascertained as common in 1 or a few populations, which can upwardly bias estimates of allele sharing among groups.[72] What proportion of variants is shared among groups if alleles are ascertained in an unbiased fashion?

Several studies have addressed this question by resequencing the regulatory and coding regions of hundreds and, more recently, thousands of genes.[32,73] Most of the common SNPs (ie, ≥10% in a given group) and haplotypes were either private (ie, population-specific) or common in only a single population, and even when common variants were shared among populations, more than 40% of these had statistically significant different allele frequencies. A comparison of the frequencies of 63 012 SNPs found in 3873 autosomal genes resequenced in individuals self-identified as Hispanic (17), African American (20), Asian American (19), and European American (20) illustrates these points. Six pairwise population comparisons of variants with frequencies of at least 10% showed that 72% to 96% of variants common in one population were present in both populations (FIGURE 3A), although only 44% to 72% of such variants were common in both populations (Figure 3B). Correlation of the frequency of common variants between populations ranged from 0.23 in African Americans vs Asian Americans to 0.83 in European Americans vs Hispanics (FIGURE 4). Correlations were consistently lower in comparisons with African Americans. Approximately 60% of variants common in at least one popu-

Figure 3. Distribution of Common Single Nucleotide Polymorphisms (SNPs) Present in Only One Population vs Common in Only One Population Among Hispanics, African Americans, Asian Americans, and European Americans



Comparison of common SNPs identified by resequencing 3873 genes in 17 Hispanics, 20 African Americans, 19 Asian Americans, and 20 European Americans. (A) The proportion of SNPs that are common (ie, ≥10%) in at least one population but found in both populations is high overall but varies from 72% to 96%. A modest proportion of common SNPs that are common in at least one population are absent in the other population. For example, only 4% of SNPs common in European Americans or Hispanics are not present in both populations, whereas 28% of SNPs common in African Americans or Asian Americans are not present in both populations. (B) The proportion of common SNPs common in both populations compared with SNPs common in only each population compared. Overall, only a modest proportion (44%-72%) of SNPs common in one population are common in both populations. A substantial proportion of common SNPs in African Americans are common only in African Americans. Data source: Genaissance Pharmaceuticals Inc, New Haven, Conn, unpublished data, August 2005.

lation were present in all populations, but only 32% of them were common in all populations.

These results indicate that a variant common in one population is, in general, present in other populations but not necessarily common in other populations. Moreover, common variants are frequently not shared between African Americans and non-African populations. Whether these findings can be generalized to the entire genome and all major human populations is unclear, but they do suggest that more data are needed before we can conclude that common variants are typically shared by all major human populations. If they are not, it might be necessary to develop initiatives to identify risk variants that are common specifically in each population studied. These findings also suggest that a sizable fraction of the genetic factors that influence common health-related traits may vary among populations. Therefore, it would be prudent to develop more initiatives to identify common variation and risk variants in non-European populations, particularly African Americans.

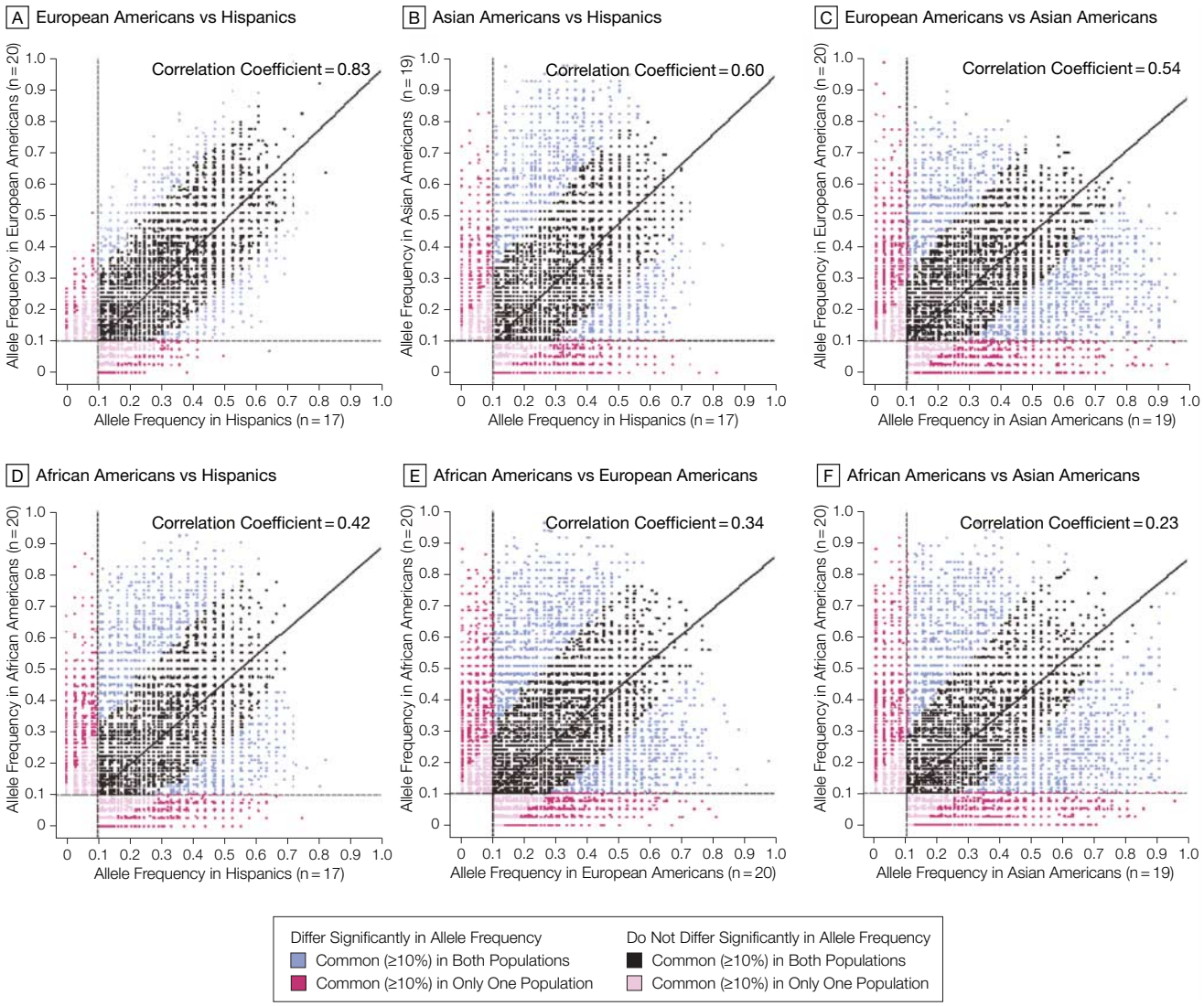
Varied Effects of Genetic Risk Factors Among Populations

Accurate inference of group membership is particularly important if genetic variants influencing common diseases have different effects in different groups. The effects of variants causing mendelian disorders and those associated with drug responses typically, but not always, have the same effect in different populations, suggesting optimistically that, in most cases, risk factors for such traits can be extrapolated across populations with different geographic ancestries. Therefore, as it becomes feasible and practical, personalized risk estimation for such health-related traits will eventually be based on direct testing for gene variants, supplanting the need to infer individual ancestry.

A recent study by Ioannidis et al[57] examined 43 gene-disease associations tested in 2 or more racial groups and for which either an overall meta-analysis showed statistically signifi-

cant results (n=32) or the results were significant in at least 1 racial group (n=11) without the meta-analysis reaching significance. These results foreshadow what we might find in general. In 25 cases (58%), the risk variant varied widely in frequency among different racial groups. Yet, for 32 of gene-disease associations, the race-specific odds ratios (ORs) were all in the same direction (eg, a gene variant that increased risk in one group also increased risk in another group). A significant association was observed in 2 racial groups in 15 meta-analyses, and in each of these cases, the effect was in the same direction. Large differences in effect were observed in only about 14% of cases. These results suggest that the effect of genetic risk factors for common disease can be extrapolated across populations. This conclusion is encouraging because had the effects of genetic risk factors frequently differed among populations, identifying variants important for the health of all humans—much less using genetics to predict individual health—could have become an intractable problem. Nevertheless, it is prudent to bear in mind

Figure 4. Comparison of Polymorphism Frequencies Among Hispanics, African Americans, Asian Americans, and European Americans



Comparison of the frequencies of single nucleotide polymorphisms (SNPs) shared among 17 Hispanics, 20 African Americans, 19 Asian Americans, and 20 European Americans in whom 3873 genes were resequenced. The less frequent SNP in the combined population was designated as the minor SNP, and the frequency of the minor SNP was calculated in each population for the 63 012 SNPs analyzed. Single nucleotide polymorphisms with significant differences ($z>1.65$, $P<.05$) between populations that are common in both populations are shown as blue data points. Single nucleotide polymorphisms with significant differences in frequency ($z>1.65$, $P<.05$) between populations that are common in only one population are shown as red dark and pale data points. Black data points represent SNPs that are common and that do not differ significantly in frequency between populations. A Spearman rank correlation coefficient between the minor allele frequencies of each SNP were estimated for each population pairwise comparison. Data source: Genaissance Pharmaceuticals Inc, New Haven, Conn, unpublished data, August 2005.

several issues that affect the overall impact of this result on understanding the importance of race on genetic factors influencing health.

First, the overwhelming majority of reported gene-disease associations have been studied in only one racial group, typically one of European ancestry. For example, of the gene-disease associations evaluated by Ioannidis et al collated from associations tested in multiple racial populations, only 48 (7%) of 667 studies were performed in populations of African origin.[57] Most gene-disease association studies do not even report the frequency of the risk variant in multiple populations. This lack of data from multiple racial groups, particularly those of African ancestry, is a major impediment to systematically determining whether the effects of disease-associated gene variants vary in effect across populations.

Second, it is unclear what proportion of overall influences on health is accounted for by genetic risk factors shared among populations. Many gene-disease associations were not included in the analysis by Ioannidis et al because the variant, disease, or both is common in only one population (eg, *TNF*α variants and cerebral malaria in sub-Saharan Africans) and hence had been studied in only one population. Such population-specific gene-disease associations might be commonplace because groups with different geographic ancestries were exposed to widely varied environments. Thus, natural selection may have increased the frequencies of risk variants in some populations but not others.[74] In 38 of the 43 meta-analyses, Ioannidis et al evaluated genetic risk factors for disease-related outcomes common mainly in older adults, usually beyond their reproductive years. The frequency of such risk variants might vary less than gene variants for health-related traits affecting mortality in children and adults during their reproductive careers. Overall, the effect of these biases would be to underestimate the effects of ancestry on the frequency of risk variants and their contribution to health-related risks.

Third, there are several examples of gene variant effects that differ among populations of different geographic ancestry. For example, an allele of apolipoprotein E (*APOE* ε*4*) that is frequent in Africans, Asians, and Europeans is associated in a dose-dependent manner with susceptibility to Alzheimer disease. However, the increased risk associated with homozygosity for *APOE* ε*4* is approximately 5-fold higher in individuals with Asian rather than African ancestry.[75] Among individuals of African ancestry, several variants in *TNF*α confer different risks for malarial disease depending on whether an individual is from West or East Africa.[50] Less frequently, risk variants have been found that are associated with ORs in different directions. One example is a combination of linked polymorphisms (ie, a haplotype) in the 5′ *cis*-regulatory region of *CCR5* that influence the rate of progression to AIDS and death.[76] Some *CCR5* haplotypes are associated with delayed disease progression in multiple racial populations, but for others the effect is population specific.[77] Moreover, one *CCR5* haplotype has been associated with delayed disease progression in European Americans with AIDS, but accelerated disease progression in African Americans.[74] Therefore, in some instances, even when the same risk variant for a complex trait is present in different racial groups, it can be associated with different effects. Predicting the effects of gene variants on disease risk and drug responses among groups will be more difficult depending on the frequency with which such phenomena occur.

There are several explanations why a variant that influences susceptibility in one population does not have the same effect in a different population. The interactions of a variant with other genes and/or with environmental factors might differ among groups. When such interactions play a role, identifying them will reduce the need for using race as a proxy to predict susceptibility. However, identifying these factors will be challenging and in the immediate future it may be appropriate to consider geographic ancestry to account for genetic and environmental correlates. A variant associated with risk might not itself be causal, but instead be in linkage disequilibrium (LD), the nonrandom association of alleles at different polymorphic sites of a chromosome, with a causal variant. If so, variation in estimated ORs might reflect the observation that LD breaks down differently in different populations. For example, Gabriel et al[72] reported an average haplotype block size of approximately 11 kb in Africans and African Americans, compared with 22 kb in European and Asian samples. Thus, variation in ORs might reflect variable LD and thus the presence of the associated variant and the causal variant on different haplotype blocks, rather than variation in a true genetic effect. How frequently this explanation is correct remains to be determined.

Conclusion

Race reflects the varied geographic ancestry of modern humans who have been partially isolated from one another throughout part of their evolutionary history. Insofar as genetic variants that influence health vary with geographic ancestry, they will also vary with race. Is race sufficient to infer ancestry to identify gene variants that predict risk of disease or drug response? In some cases, its accuracy may be adequate, but the information about genetic group membership captured by notions of race is, in general, less than that obtained by making inferences of ancestry from geographic or explicit genetic data. For example, to identify risk factors for disease and predictors of treatment response that influence a person's health, it is likely less useful to know that a person is Hispanic than whether a person's ancestors came from a country such as Puerto Rico, where there is substantial African admixture, or one in which there was a great deal of historical admixture with Native Americans such as Mexico. Therefore, in circumstances in which ancestry is predictive of genetic risk factors asso-

ciated with health-related traits, using race rather than geography or explicit genetic data to infer ancestry will be less useful for making decisions about disease risk or treatment response.

Much remains to be learned about the relationship between race, ancestry, and health. Genetic inference of geographic ancestry will improve as human genetic diversity is assessed at an increasingly finer spatial resolution. Nevertheless, genetic inference of geographic ancestry is likely to remain imperfect at best. However, making use of individual genetic profiles to promote a healthier life by means of targeted prevention (eg, use of higher-dose folic acid supplements in women homozygous for C677T in *MTHFR*) and intervention (eg, use of an analgesic other than codeine in the 10% of individuals of European ancestry who experience no analgesic effect because of null alleles in *CYP2D6*) strategies may not require more than the rough approximation of geographic ancestry afforded by genetic inference. In what circumstances such inferences work needs to be addressed empirically. However, genes interact with the environment and with each other in complex ways. Sorting out this complexity will require balanced, interdisciplinary approaches that do not take away from identifying and rectifying the environmental causes of health disparities among groups.

Empirical data on the correspondence of race with health-related traits are still too limited to draw definitive conclusions. Indeed, the paucity of data on gene-disease associations in individuals of African ancestry is disturbing. The extent to which gene variants underlying common diseases differ in frequency and/or effect between groups can be addressed only if data are available from well-designed studies in multiple populations. Even so, because of substantial variation within human populations, it is certain that labels such as race will often be an inaccurate proxy when making decisions about disease predisposition, drug response, and the like.

Attributing racial labels to DNA samples is well on its way to acceptance in forensics and law enforcement.[78] Application in medicine is likely forthcoming, but has arguably greater potential to cause harm, as misuse of genetic information about race in medical applications could severely undermine the public's confidence in the application of genetics to health. Referring to "geographic ancestry" instead of race is an emerging alternative that is both more accurate and less contentious. One way to operationalize this approach is for the National Institutes of Health to change its current requirement to use Office of Management and Budget categories and instead mandate stratification of individuals by self-assessed descriptors of ancestry such as the geographic origin of an individual's parents (eg, Central Africa, Southeast Asia, Central America), followed by their ethnic identity, and finally the community in which a person resides. This strategy might not differ much from existing practices—particularly for individuals who know little about their origins—but it underscores the need to take account of biogeographic ancestry, it de-emphasizes the use of racial categories, and it may be a better interim solution to making ancestry inferences in the absence of explicit genetic data. In the end, however, every human being is genetically unique and so must be treated as an individual, not an example of a group defined by geography or race.

Financial Disclosures: None reported.
Funding/Support: This work was supported by grants HD48895, AI46326, AI65357, and RR00064 from the National Institutes of Health; grant U50/CCU822097 from the Centers for Disease Control and Prevention; and an unrestricted grant from the Primary Children's Medical Center Foundation, Salt Lake City, Utah.
Role of the Sponsors: The funding sources had no role in the development or synthesis of this review or in the writing of the manuscript.
Acknowledgment: I thank the following individuals for their assistance: Steve Guthery, MD, Department of Pediatrics (statistical analysis and discussion), Stephen Wooding, PhD, Department of Human Genetics (discussion), and Lynn Jorde, PhD, Department of Human Genetics (discussion), University of Utah, Salt Lake City; Ben Salisbury, PhD, Genaissance Pharmaceuticals, New Haven, Conn (unpublished data and discussion).

REFERENCES

1. Osborne NG, Feit MD. The use of race in medical research. *JAMA*. 1992;267:275-279.
2. Oppenheimer G. Paradigm lost: race, ethnicity, and the search for a new population taxonomy. *Am J Public Health*. 2001;91:1049-1055.
3. Institute of Medicine. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: National Academies Press; 2003.
4. Risch N, Burchard E, Ziv E, Tang H. Categorization of humans in biomedical research: genes, race, and disease. *Genome Biol*. 2002;3:1-12.
5. Burchard EG, Ziv E, Coyle N, et al. The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med*. 2003;348:1170-1175.
6. Cooper RS, Kaufman JS, Ward R. Race and genomics. *N Engl J Med*. 2003;348:1166-1170.
7. Wilson JF, Weale ME, Smith AC, et al. Population genetic structure of variable drug response. *Nat Genet*. 2001;29:265-269.
8. Bamshad MJ, Olson SE. Does race exist? *Sci Am*. 2003;289:78-85.
9. Sehgal AR. Overlap between whites and blacks in response to antihypertensive drugs. *Hypertension*. 2004;43:566-571.
10. Sethi AA, Nordestgaard BG, Tybjaerg-Hansen A. Angiotensinogen gene polymorphism, plasma angiotensinogen, and risk of hypertension and ischemic heart disease: a meta-analysis. *Arterioscler Thromb Vasc Biol*. 2003;23:1269-1275.
11. Rosskopf D, Manthey I, Siffert W. Identification and ethnic distribution of major haplotypes in the gene GNB3 encoding the G-protein beta3 subunit. *Pharmacogenetics*. 2002;12:209-220.
12. Yen CJ, Beamer BA, Negri C, et al. Molecular scanning of the human peroxisome proliferator activated receptor gamma (hPPAR gamma) gene in diabetic Caucasians: identification of a Pro12Ala PPAR gamma 2 missense mutation. *Biochem Biophys Res Commun*. 1997;241:270-274.
13. Ezenwaka C, Kalloo R, Uhlig M, et al. The E23K variant in the Kir6.2 subunit of the ATP-sensitive K+ channel does not augment impaired glucose tolerance in Caribbean subjects with a family history of type 2 diabetes. *J Endocrinol*. 2005;185:439-444.
14. Riedel MJ, Steckley DC, Light PE. Current status of the E23K Kir6.2 polymorphism: implications for type-2 diabetes. *Hum Genet*. 2005;116:133-145.
15. Cohen J, Pertsemlidis A, Kotowski IK, et al. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet*. 2005;37:161-165.
16. Mori M, Yamada R, Kobayashi K, et al. Ethnic differences in allele frequency of autoimmune-disease-associated SNPs. *J Hum Genet*. 2005;50:264-266.
17. Miller LH, Mason SJ, Dvorak JA, et al. Erythrocyte receptors for (*Plasmodium knowlesi*) malaria: Duffy blood group determinants. *Science*. 1975;189:561-563.
18. Xie HG, Kim RB, Wood AJ, et al. Molecular basis of ethnic differences in drug disposition and response. *Annu Rev Pharmacol Toxicol*. 2001;41:815-850.
19. Zhu X, Luke A, Cooper RS, et al. Admixture mapping for hypertension loci with genome-scan markers. *Nat Genet*. 2005;37:177-181.
20. Fernandez JR, Shriver MD, Beasley TM, et al. Association of African genetic admixture with resting metabolic rate and obesity among women. *Obes Res*. 2003;11:904-911.
21. Gower BA, Fernandez JR, Beasley TM, et al. Using genetic admixture to explain racial differences in insulin-related phenotypes. *Diabetes*. 2003;52:1047-1051.
22. Sankar P, Cho MK, Condit CM, et al. Genetic research and health disparities. *JAMA*. 2004;291:2985-2989.
23. Cooper RS. Social inequality, ethnicity, and cardiovascular disease. *Int J Epidemiol*. 2001;30:S48-S52.
24. Bradley EH, Herrin J, Wang Y, et al. Racial and ethnic differences in time to acute reperfusion therapy

for patients hospitalized with myocardial infarction. *JAMA*. 2004;292:1563-1572.
25. Li WH, Sadler LA. Low nucleotide diversity in man. *Genetics*. 1991;129:513-523.
26. Kruglyak L, Nickerson DA. Variation is the spice of life. *Nat Genet*. 2001;27:234-236.
27. Fay JC, Wyckoff GJ, Wu CI. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature*. 2002;415:1024-1026.
28. Harpending H, Rogers AR. Genetic perspectives on human origins and differentiation. *Annu Rev Genomics Hum Genet*. 2000;1:361-385.
29. Harding RM, McVean G. A structured ancestral population for the evolution of modern humans. *Curr Opin Genet Dev*. 2004;14:667-674.
30. Rosenberg NA, Pritchard JK, Weber JL, et al. Genetic structure of human populations. *Science*. 2002; 298:2381-2385.
31. Bamshad MJ, Wooding SW, Watkins WS, et al. Human population genetic structure and inference of group membership. *Am J Hum Genet*. 2003;72:578-589.
32. Bamshad M, Wooding SW, Salisbury BA, Stephens JC. Deconstructing the relationship between genetics and race. *Nat Rev Genet*. 2004;5:598-608.
33. Jorde LB, Watkins WS, Bamshad MJ, et al. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet*. 2000;66:979-988.
34. Edwards AWF. Human genetic diversity: Lewontin's fallacy. *Bioessays*. 2003;25:798-801.
35. Shriver MD, Mei R, Parra EJ, et al. Large-scale SNP analysis reveals clustered and continuous patterns of human genetic variation. *Hum Genomics*. 2005;2:81-89.
36. Rosenberg N, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet*. 2003;73:1402-1422.
37. Tang H, Quertermous T, Rodriguez B, et al. Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *Am J Hum Genet*. 2005;76:268-275.
38. Tate SK, Goldstein DB. Will tomorrow's medicines work for everyone? *Nat Genet*. 2004;36:S34-S42.
39. Haga SB, Venter JC. FDA races in wrong direction. *Science*. 2003;301:466.
40. Kittles RA, Weiss KM. Race, ancestry, and genes: implications for defining disease risk. *Annu Rev Genomics Hum Genet*. 2003;4:33-67.
41. United States Census. General Demographic Characteristics (DP-1) 2000. Available at: http://factfinder.census.gov/servlet/QTTable?_bm=y&-geo_id=01000US&-qr_name=DEC_2000_SF1_U_DP1&-ds_name=DEC_2000_SF1_U. Accessibility verified July 29, 2005.
42. Shriver MD, Parra EJ, Dios S, et al. Skin pigmentation, biogeographical ancestry, and admixture mapping. *Hum Genet*. 2003;112:387-399.
43. Reiner AP, Ziv E, Lind DL, et al. Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. *Am J Hum Genet*. 2005;76:463-477.
44. Frudakis T, Venkateswarlu K, Thomas MJ, et al. A classifier for the SNP-based inference of ancestry. *J Forensic Sci*. 2003;48:771-782.
45. Abe-Sandes K, Silva WA, Zago MA. Heterogeneity of the Y chromosome in Afro-Brazilian populations. *Hum Biol*. 2004;76:77-86.
46. Tishkoff SA, Williams SM. Genetic analysis of African populations: human evolution and complex disease. *Nat Rev Genet*. 2002;3:611-621.
47. Satta Y, Takahata N. The distribution of the ancestral haplotype in finite stepping-stone models with population expansion. *Mol Ecol*. 2004;13:877-886.
48. Watkins WS, Rogers AR, Ostler CT, et al. Genetic variation among world populations: inferences from 100 *Alu* insertion polymorphisms. *Genome Res*. 2003;13:1607-1618.
49. Yu N, Chen FC, Ota S, et al. Larger genetic differences within Africans than between Africans and Eurasians. *Genetics*. 2002;161:269-274.
50. Knight JC, Udalova I, Hill AV, et al. A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. *Nat Genet*. 1999;22:145-150.
51. Ptak SE, Przeworski M. Evidence for population growth in humans is confounded by fine-scale population structure. *Trends Genet*. 2002;18:559-563.
52. Hinds DA, Stuve LL, Nilsen GB, et al. Whole-genome patterns of common DNA variation in three human populations. *Science*. 2005;307:1072-1079.
53. The international HapMap Consortium. The international HapMap Project. *Nature*. 2003;426:789-796.
54. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant . . . or not? *Hum Mol Genet*. 2002;11:2417-2423.
55. Badano JL, Katsanis N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nat Rev Genet*. 2002;3:779-789.
56. Lohmueller KE, Pearce CL, Pike M, et al. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet*. 2003;33:177-182.
57. Ioannidis JPA, Ntzani EE, Trikalinos TA. "Racial" differences in genetic effects for complex diseases. *Nat Genet*. 2004;36:1312-1318.
58. Ackerman MJ, Tester DJ, Jones GS, et al. Ethnic differences in cardiac potassium channel variants: implications for genetic susceptibility to sudden cardiac death and genetic testing for congenital long QT syndrome. *Mayo Clin Proc*. 2003;78:1479-1487.
59. Basehore MJ, Howard TD, Lange LA, et al. A comprehensive evaluation of IL4 variants in ethnically diverse populations: association of total serum IgE levels and asthma in white subjects. *J Allergy Clin Immunol*. 2004;114:80-87.
60. Pang CP, Lam DS. Differential occurrence of mutations causative of eye diseases in the Chinese population. *Hum Mutat*. 2002;19:189-208.
61. Frodsham AJ, Hill AV. Genetics of infectious diseases. *Hum Mol Genet*. 2004;13:187-194.
62. Gonzalez E, Dhanda R, Bamshad M, et al. Global survey of genetic variation in CCR5, RANTES, and MIP-1α: impact on the epidemiology of the HIV-1 pandemic. *Proc Natl Acad Sci U S A*. 2001;98:5199-5204.
63. Siffert W, Forster P, Jockel KH, et al. Worldwide ethnic distribution of the G protein beta3 subunit 825T allele and its association with obesity in Caucasian, Chinese, and Black African individuals. *J Am Soc Nephrol*. 1999;10:1921-1930.
64. Ogura Y, Bonen DK, Inohara N, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*. 2001;411:603-606.
65. Begovich AB, Carlton VE, Honigberg LA, et al. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet*. 2004;75:330-337.
66. Kyogoku C, Langefeld CD, Ortmann WA, et al. Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. *Am J Hum Genet*. 2004;75:504-507.
67. Kugathasan S, Loizides A, Babusukumar U, et al. Comparative phenotypic and CARD15 mutational analysis among African American, Hispanic, and white children with Crohn's disease. *Inflamm Bowel Dis*. 2005;11:631-638.
68. Gregersen PK. Pathways to gene identification in rheumatoid arthritis: PTPN22 and beyond. *Immunol Rev*. 2005;204:74-86.
69. Gonzalez E, Kulkarni H, Bolivar H, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science*. 2005; 307:1434-1440.
70. Thomas PD, Kejariwal A. Coding single-nucleotide polymorphisms associated with complex vs Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci U S A*. 2004;101:15398-15403.
71. Zhang J, Rowe WL, Clark AG, Buetow KH. Genomewide distribution of high-frequency, completely mismatching SNP haplotype pairs observed to be common across human populations. *Am J Hum Genet*. 2003;73:1073-1081.
72. Gabriel SB, Schaffner SD, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science*. 2002;296:2225-2229.
73. Carlson CS, Eberle MA, Rieder MJ, et al. Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nat Genet*. 2003;33:518-521.
74. Bamshad M, Wooding SP. Signatures of natural selection in the human genome. *Nat Rev Genet*. 2003; 4:99-111.
75. Farrer LA, Cupples LA, Haines JL, et al; APOE and Alzheimer Disease Meta Analysis Consortium. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis. *JAMA*. 1997;278:1349-1356.
76. Martin MP, Dean M, Smith MW, et al. Genetic acceleration of AIDS progression by a promoter variant of *CCR5*. *Science*. 1998;282:1907-1911.
77. Gonzalez E, Bamshad M, Sato S, et al. Race-specific HIV-1 disease-modifying effects associated with *CCR5* haplotypes. *Proc Natl Acad Sci U S A*. 1999;96: 12004-12009.
78. Cho MK, Sankar P. Forensic genetics and ethical, legal, and social implications beyond the clinic. *Nat Genet*. 2004;36:S8-S12.

Luiz A. R. De Freitas, MD
CPqGM (Gonçalo Muniz Research Center)–FIOCRUZ (Osvaldo Cruz Foundation)
Bahia, Brazil

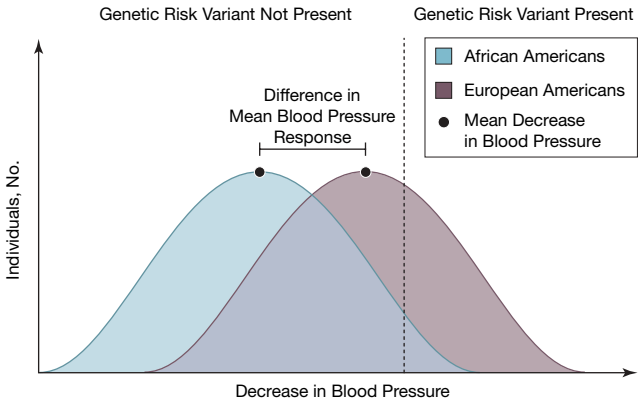
Author Contributions: Dr Cotrim had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Cotrim, Carvalho, De Freitas.
Acquisition of data: Cotrim, Siqueira, Lordelo, Rocha, De Freitas.
Analysis and interpretation of data: Cotrim, Carvalho, De Freitas.
Drafting of the manuscript: Carvalho, Siqueira, Lordelo, Rocha.
Critical revision of the manuscript for important intellectual content: Cotrim, Carvalho, De Freitas.
Statistical analysis: Carvalho, Siqueira, Lordelo, Rocha.
Obtained funding: Cotrim, Carvalho.
Administrative, technical, or material support: Cotrim, Carvalho, De Freitas.
Study supervision: Cotrim, Carvalho.
Financial Disclosures: None reported.
Funding Support: This work was supported by CEPETRO-CNPq (Brazilian Government National Research Council) research grant 464438/00-4 and FINEP (Brazilian Government Research Funding Foundation) research grant FINEP/CTPETRO 64.000.350.00.
Role of the Sponsors: The sponsors had no role in the design and conduct of the study; in the collection, management, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript.

1. Marchesini G, Burganesi E, Forlani G, et al. Nonalcoholic fatty liver, steatohepatitis and metabolic syndrome. *Hepatology*. 2003;37:917-923.
2. Neuschwander-Tetri BA, Caldwell SH. Nonalcoholic steatohepatitis: summary of an AASLD Single Topic Conference. *Hepatology*. 2003;37:1202-1219.
3. Cotrim HP, Andrade ZA, Parana R, Portugal M, Lyra LG, Freitas LA. Nonalcoholic steatohepatitis: a toxic liver disease in industrial workers. *Liver*. 1999;19:299-304.
4. Cotrim HP, De Freitas LA, Freitas C, et al. Clinical and histopathological features of NASH in workers exposed to chemicals with or without associated metabolic conditions. *Liver Int*. 2004;24:131-135.
5. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*. 1985;28:412-419.
6. Foss-Freitas MC, Foss MC. Comparison of the homeostasis model assessment and quantitative insulin sensitivity check index with data from forearm metabolic studies for the in vivo assessment of insulin sensitivity. *Braz J Med Biol Res*. 2004;37:663-668.
7. Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, McCullough AJ. Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity. *Gastroenterology*. 1999;116:1413-1419.

CORRECTION

Incorrect Figure: In the Special Communication entitled "Genetic Influences on Health: Does Race Matter?" published in the August 24/31, 2005, issue of *JAMA* (2005;294:937-946), there is an error in **FIGURE 1**. The colors for the distributions below the curve were reversed; the key to the Figure is correct. The correct Figure 1 appears below.

Figure 1. Hypothetical Relationship Between Genetic Risk, Ancestry, and Race



Distributions of the reduction in blood pressure observed in African Americans and European Americans after treatment with an angiotensin-converting enzyme (ACE) inhibitor. One hypothetical explanation for the mean difference in treatment response is that a genetic risk variant predictive of a positive response to treatment is more common in European Americans (individuals to the right of the dotted line) than in African Americans. Note, however, that some African Americans also have the genetic risk variant and that many African Americans and European Americans who do not have the genetic risk variant have a similar response to treatment (ie, overlap between distributions). In this case, race might not be considered a good predictor of genetic risk or response to treatment. Based on an original concept by Seghal.[9]